Check for updates

Pilot study of freshly excised breast tissue response in the 300–600 GHz range

QUENTIN CASSAR,[1] AMEL AL-IBADI,[1] LAVEN MAVARANI,[2] PHILIPP HILLGER,[2] JANUSZ GRZYB,[2] GAËTAN MACGROGAN,[3] THOMAS ZIMMER,[1] ULLRICH R. PFEIFFER,[2] JEAN-PAUL GUILLET,[1] AND PATRICK MOUNAIX[1,*]

[1]*Integration: from Material to Systems Laboratory, UMR CNRS 5218, University of Bordeaux, 33400 Talence, France*
[2]*Institute for High-Frequency and Communication Technology, University of Wuppertal, 42119 Wuppertal, Germany*
[3]*Department of Pathology, Bergonié Institute, 33076 Bordeaux, France*
*patrick.mounaix@u-bordeaux.fr

Abstract: The failure to accurately define tumor margins during breast conserving surgery (BCS) results in a 20% re-excision rate. The present paper reports the investigation to evaluate the potential of terahertz imaging for breast tissue recognition within the under-explored 300–600 GHz range. Such a frequency window matches new BiCMOS technology capabilities and thus opens up the opportunity for near-field terahertz imaging using these devices. To assess the efficacy of this frequency band, data from 16 freshly excised breast tissue samples were collected and analyzed directly after excision. Complex refractive indices have been extracted over the as-mentioned frequency band, and amplitude frequency images show some contrast between tissue types. Principal component analysis (PCA) has also been applied to the data in an attempt to automate tissue classification. Our observations suggest that the dielectric response could potentially provide contrast for breast tissue recognition within the 300–600 GHz range. These results open the way for silicon-based terahertz subwavelength near field imager design, efficient up to 600 GHz to address ex vivo life-science applications.

References and links

1. J. P. Guillet, B. Recur, L. Frederique, B. Bousquet, L. Canioni, I. Manek-Hönninger, P. Desbarats, and P. Mounaix, "Review of terahertz tomography techniques," J. Infrared Millim. Terahertz Waves **35**(4), 382–411 (2014).
2. C. Roman, O. Ichim, L. Sarger, V. Vigneras, and P. Mounaix, "Terahertz dielectric characterisation of polymethacrylimide rigid foam: The perfect sheer plate?" Electron. Lett. **40**(19), 1167 (2004).
3. J. Bou-Sleiman, J.-B. J.-B. Perraud, B. Bousquet, J.-P. J.-P. Guillet, N. Palka, and P. Mounaix, "Discrimination and identification of RDX/PETN explosives by chemometrics applied to terahertz time-domain spectral imaging," in *Millimetre Wave and Terahertz Sensors and Technology VIII* (International Society for Optics and Photonics, 2015), Vol. **9651**, p. 965109.
4. C. Jansen, S. Wietzke, O. Peters, M. Scheller, N. Vieweg, M. Salhi, N. Krumbholz, C. Jördens, T. Hochrein, and M. Koch, "Terahertz imaging: applications and perspectives," Appl. Opt. **49**(19), E48–E57 (2010).
5. Y. C. Shen, "Terahertz pulsed spectroscopy and imaging for pharmaceutical applications: A review," Int. J. Pharm. **417**(1-2), 48–60 (2011).
6. P. F. Taday, "Applications of terahertz spectroscopy to pharmaceutical sciences," Philos Trans A Math Phys Eng Sci **362**(1815), 351–364, discussion 363–364 (2004).
7. P. H. Siegel, "Terahertz technology in biology and medicine," IEEE Trans. Microw. Theory Tech. **52**(10), 2438–2447 (2004).
8. E. Pickwell and V. P. Wallace, "Biomedical applications of terahertz technology," J. Phys. D Appl. Phys. **39**(17), R301–R310 (2006).
9. G. S. Park, Y. H. Kim, H. Han, J. K. Han, J. Ahn, J. H. Son, W. Y. Park, and Y. U. Jeong, *Convergence of Terahertz Sciences in Biomedical Systems* (Springer, 2012).
10. S. Fan, Y. He, B. S. Ung, and E. Pickwell-Macpherson, "The growth of biomedical terahertz research," J. Phys. D Appl. Phys. **47**(37), 374009 (2014).

11. P. H. Siegel and V. Pikov, "Impact of low intensity millimetre waves on cell functions," Electron. Lett. **46**(26), S70 (2010).

12. S. Sy, S. Huang, Y.-X. J. X. J. Wang, J. Yu, A. T. Ahuja, Y. T. Zhang, and E. Pickwell-MacPherson, "Terahertz spectroscopy of liver cirrhosis: investigating the origin of contrast," Phys. Med. Biol. **55**(24), 7587–7596 (2010).

13. M. H. Arbab, D. P. Winebrenner, T. C. Dickey, A. Chen, M. B. Klein, and P. D. Mourad, "Terahertz spectroscopy for the assessment of burn injuries in vivo," J. Biomed. Opt. **18**(7), 077004 (2013).

14. Z. D. Taylor, R. S. Singh, D. B. Bennett, P. Tewari, C. P. Kealey, N. Bajwa, M. O. Culjat, A. Stojadinovic, H. Lee, J.-P. P. Hubschman, E. R. Brown, and W. S. Grundfest, "THz medical imaging: in vivo hydration sensing," IEEE Trans. Terahertz Sci. Technol. **1**(1), 201–219 (2011).

15. Y. Miura, A. Kamataki, M. Uzuki, T. Sasaki, J. Nishizawa, and T. Sawai, "Terahertz-wave spectroscopy for precise histopathological imaging of tumor and non-tumor lesions in paraffin sections," Tohoku J. Exp. Med. **223**(4), 291–296 (2011).

16. P. Doradla, K. Alavi, C. Joseph, and R. Giles, "Single-channel prototype terahertz endoscopic system," J. Biomed. Opt. **19**(8), 080501 (2014).

17. S. J. Oh, S.-H. Kim, Y. B. Ji, K. Jeong, Y. Park, J. Yang, D. W. Park, S. K. Noh, S.-G. Kang, Y.-M. Huh, J.-H. Son, and J.-S. Suh, "Study of freshly excised brain tissues using terahertz imaging," Biomed. Opt. Express **5**(8), 2837–2842 (2014).

18. C. S. Joseph, A. N. Yaroslavsky, V. A. Neel, T. M. Goyette, and R. H. Giles, "Continuous wave terahertz transmission imaging of nonmelanoma skin cancers," Lasers Surg. Med. **43**(6), 457–462 (2011).

19. R. M. Woodward, B. E. Cole, V. P. Wallace, R. J. Pye, D. D. Arnone, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue," Phys. Med. Biol. **47**(21), 3853–3863 (2002).

20. P. C. Ashworth, E. Pickwell-MacPherson, E. Provenzano, S. E. Pinder, A. D. Purushotham, M. Pepper, and V. P. Wallace, "Terahertz pulsed spectroscopy of freshly excised human breast cancer," Opt. Express **17**(15), 12444–12454 (2009).

21. A. M. Hassan, D. C. D. C. Hufnagle, M. El-Shenawee, and G. E. G. E. Pacey, "Terahertz imaging for margin assessment of breast cancer tumors," in *Microwave Symposium Digest (MTT), 2012 IEEE MTT-S International* (IEEE, 2012), **Vol. 12**, pp. 1–3.

22. A. J. Fitzgerald, S. Pinder, A. D. Purushotham, P. O'Kelly, P. C. Ashworth, and V. P. Wallace, "Classification of terahertz-pulsed imaging data from excised breast tissue," J. Biomed. Opt. **17**(1), 016005 (2012).

23. J. Landercasper, E. Whitacre, A. C. Degnim, and M. Al-Hamadani, "Reasons for re-excision after lumpectomy for breast cancer: insight from the american society of breast surgeons mastery(SM) database," Ann. Surg. Oncol. **21**(10), 3185–3191 (2014).

24. R. Jeevan, D. A. Cromwell, M. Trivella, G. Lawrence, O. Kearins, J. Pereira, C. Sheppard, C. M. Caddy, and J. H. P. van der Meulen, "Reoperation rates after breast conserving surgery for breast cancer among women in England: retrospective study of hospital episode statistics," BMJ **345**(jul12 2), e4505 (2012).

25. M. R. Grootendorst, A. J. Fitzgerald, S. G. Brouwer de Koning, A. Santaolalla, A. Portieri, M. Van Hemelrijck, M. R. Young, J. Owen, M. Cariati, M. Pepper, V. P. Wallace, S. E. Pinder, and A. Purushotham, "Use of a handheld terahertz pulsed imaging device to differentiate benign and malignant breast tissue," Biomed. Opt. Express **8**(6), 2932–2945 (2017).

26. A. J. Fitzgerald, V. P. Wallace, M. Jimenez-Linan, L. Bobrow, R. J. Pye, A. D. Purushotham, and D. D. Arnone, "Terahertz Pulsed Imaging of human breast tumors," Radiology **239**(2), 533–540 (2006).

27. T. Bowman, Y. Wu, J. Gauch, L. K. Campbell, and M. El-Shenawee, "Terahertz Imaging of Three-Dimensional Dehydrated Breast Cancer Tumors," J. Infrared Millim. Terahertz Waves **38**(6), 766–786 (2017).

28. V. P. Wallace, A. J. Fitzgerald, E. Pickwell, R. J. Pye, P. F. Taday, N. Flanagan, and T. Ha, "Terahertz pulsed spectroscopy of human Basal cell carcinoma," Appl. Spectrosc. **60**(10), 1127–1133 (2006).

29. T. Bowman, T. Chavez, K. Khan, J. Wu, A. Chakraborty, N. Rajaram, K. Bailey, and M. El-Shenawee, "Pulsed terahertz imaging of breast cancer in freshly excised murine tumors," J. Biomed. Opt. **23**(2), 1–13 (2018).

30. A. Al-Ibadi, Q. Cassar, T. Zimmer, G. MacGrogan, L. Mavarani, P. Hillger, J. Grzyb, U. R. Pfeiffer, J. P. Guillet, and P. Mounaix, "THz spectroscopy and imaging for breast cancer detection in the 300-500 GHz range," in *International Conference on Infrared, Millimeter, and Terahertz Waves, IRMMW-THz* (2017).

31. R. C. Gonzalez and R. E. Woods, "Digital Image Processing (3rd Edition)," 3Rd Ed. 954 (2008).

32. L. Mavarani, P. Hillger, J. Grzyb, Q. Cassar, A. Al-Ibadi, T. Zimmer, G. MacGrogan, J. P. P. Guillet, P. Mounaix, and U. R. R. Pfeiffer, "A novel approach of aqueous solution analysis using a fully-integrated terahertz near-field sensor," in *Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), 2017 42nd International Conference on* (IEEE, 2017), pp. 1–2.

33. J. Grzyb, B. Heinemann, and U. R. Pfeiffer, "Solid-State Terahertz Superresolution Imaging Device in 130-nm SiGe BiCMOS Technology," IEEE Trans. Microw. Theory Tech. **65**(11), 4357–4372 (2017).

34. P. Hillger, R. Jain, J. Grzyb, L. Mavarani, B. Heinemann, G. MacGrogan, P. Mounaix, T. Zimmer, and U. Pfeiffer, "A 128-pixel 0.56THz sensing array for real-time near-field imaging in 0.13μm SiGe BiCMOS," in *IEEE International Solid - State Circuits Conference - (ISSCC)* (2018).

35. T. Löffler, K. Siebert, S. Czasch, T. Bauer, and H. G. Roskos, "Visualization and classification in biomedical terahertz pulsed imaging," Phys. Med. Biol. **47**(21), 3847–3852 (2002).

36. H. Balacey, B. Recur, J.-B. J.-B. B. Perraud, J. B. J. B. J. B. Sleiman, J.-P. J.-P. P. J.-P. Guillet, and P. Mounaix, "Advanced Processing Sequence for 3-D THz Imaging," IEEE Trans. Terahertz Sci. Technol. **6**(2), 191–198

(2016).

37. H. Ballacey, A. Al-Ibadi, G. Macgrogan, J. P. P. Guillet, E. Macpherson, and P. Mounaix, "Automated data and image processing for biomedical sample analysis," in *Infrared, Millimeter, and Terahertz Waves (IRMMW-THz), 2016 41st International Conference on* (IEEE, 2016), Vol. 2016–Novem, pp. 1–2.

38. H. Abdi and L. J. Williams, "Principal component analysis," Wiley Interdisc. Rev. Comput. Stat. **2**(4), 433–459 (2010).

39. S. Fan, E. P. J. Parrott, B. S. Y. Ung, and E. Pickwell-MacPherson, "Calibration method to improve the accuracy of THz imaging and spectroscopy in reflection geometry," Photon. Res. **4**(3), A29–A35 (2016).

40. P. U. Jepsen, U. Møller, and H. Merbold, "Investigation of aqueous alcohol and sugar solutions with reflection terahertz time-domain spectroscopy," Opt. Express **15**(22), 14717–14737 (2007).

41. A. Pashkin, M. Kempa, H. Němec, F. Kadlec, and P. Kužel, "Phase-sensitive time-domain terahertz reflection spectroscopy," Rev. Sci. Instrum. **74**(11), 4711–4717 (2003).

42. M. Koeberg, C.-C. Wu, D. Kim, and M. Bonn, "THz dielectric relaxation of ionic liquid : water mixtures," Chem. Phys. Lett. **439**(1-3), 60–64 (2007).

43. A. J. Fitzgerald, E. Pickwell-MacPherson, and V. P. Wallace, "Use of finite difference time domain simulations and Debye theory for modelling the terahertz reflection response of normal and tumour breast tissue," PLoS One **9**(7), e99291 (2014).

44. E. Pickwell, B. E. Cole, A. J. Fitzgerald, V. P. Wallace, and M. Pepper, "Simulation of terahertz pulse propagation in biological systems," Appl. Phys. Lett. **84**(12), 2190–2192 (2004).

45. M. Scheller, S. Wietzke, C. Jansen, and M. Koch, "Modelling heterogeneous dielectric mixtures in the terahertz regime: A quasi-static effective medium theory," J. Phys. D Appl. Phys. **42**(6), 065415 (2009).

1. Introduction

Ranging between far infrared and microwaves, the development of terahertz science and technology comes from both the electronic and optic sides and thus terahertz (THz) technology takes advantages from these two fields. Consequently, the catalog of applications has increased and includes for example quality control, non-destructive testing for 3D imaging [1], spectroscopic characterization of materials [2] and chemical recognition [3]. Many of these applications rely on the unique spectral fingerprints, in the terahertz range, of chemicals and materials. This key fact can be exploited in many industrial applications [4] especially for the pharmaceutical and chemical industries [5,6].

Also, the potential of terahertz time-domain spectroscopy (THz-TDS) has opened new horizons in the field of medicine and biology [7–10]. One of the benefits to use THz-TDS for tissue examination is that the low energy of THz photons is expected not to modify the global biochemical structure and thus the basic cell functions are assumed to be conserved through analysis [11]. THz-TDS has been successively studied for various medical applications like liver cirrhosis [12], burn injuries [13,14] or myocardial infarction [15]. Cancer diagnosis has also been investigated through the use of THz-TDS [16–19], notably for breast cancer [20–22]. Breast cancer is one of the most common diseases among women and is proving to be particularly invasive. As a results of technological advances toward early breast cancer detection, segmental mastectomy is increasingly ordinary, limiting total breast removal rate. However, the precision with which are delineated tumor margins remains weak and periodically leads to a second surgery to assess the entire breast cancer removal. In fact, following histopathologic examination of excised tissues, up to 20% are shown to have tumor at the margins [23,24]. A second surgical procedure is then required to minimize the recurrence risk but results in additional cost and increased morbidity rate. Consequently there is a clear need for an operating surgery room device which could accurately define breast tumor margins during breast conserving surgeries [25]. *Fitzgerald et al* [26], were the first to investigate the ability of THz-TDS to map the breast tumor margins. Spectro-imaging measurements on paraffin embedded and fresh tissues [20,27] having highlighted that both the refractive index and the absorption coefficient of basal cell carcinoma are higher than the adjacent normal tissue ones. Highest differences between tissue types have been reported around 500 GHz [20]. It has been suggested that the substantial contrast between normal and abnormal tissues arises from the different water content to which THz-photons are sensitive [28]. Worldwide, the accumulation of data is expanding to certify the ability of the THz range to provide tissue recognition based on dielectric properties [29,30]. Moreover, image

processing techniques are studied and developed to improve the visualization of THz-images [31].

In an attempt to develop a new way to detect tumor margins, we have investigated the responses of freshly excised breast tissue through 16 different samples in the 300 – 600 GHz band, using a THz-TDS system situated in a hospital. The objective is to analyze the tissue type responses to submillimeter frequencies that match with BiCMOS technology capabilities [32,33]. Positive results would lead to the design of a new terahertz near-field imaging array working within this frequency band [34]. Indeed, near-field analysis would provide a resolution closer to the typical eukaryote cell diameter and thus, is expected to enhance the accuracy of tissue recognition. Moreover, this new technology should be more sensitive to permittivity changes, smaller, faster and cheaper, overcoming the main THz-TDS issues.

Complementary to an automatic signal and statistical data processing [35] we developed for medical applications [36,37], PCA [38] within the 300 – 600 GHz band has been conducted to explore its aptitude to manage tissue recognition. Processed data provide motivating results that could be used to assess an automatic tissue classification and interface delimitation tool.

2. Methods

2.1 Acquisition set-up

A commercially available TeraPulse 4000 (Teraview Ltd, Cambridge, UK), with a modified reflection geometry shown in Fig. 1, was used in this study. THz pulses are generated by focusing ultrafast near-infrared pulses onto the gap between two electrodes of a GaAs photo-switch. The acceleration of the free carriers creates a transient photocurrent and coupled to an antenna, generates an electromagnetic field with a bandwidth lying between 100 GHz and 4 THz. The emitted THz pulses are then focused, using a 50 mm focal length plano-convex tsurupica lens (Microtech Instruments, USA), on a 2 mm thick C-cut sapphire substrate. The C-cut sapphire window has been selected to avoid birefringence effect. The sapphire substrate is mounted on a motor-stage to perform reflection THz-imaging. The step sizes in X and Y directions was either 100 μm or 500 μm depending on the size of the sample and to limit the acquisition time. Coherent photoconductive detection of the reflected pulses is performed using a similar photoconductive antenna as used for emission. The experimental atmosphere is purged of water molecules through dried-air injection. Thus a signal-to-noise ratio (SNR) of 55 dB can be achieved on an air/metal dielectric interface. The spatial resolution (SR) achieved using our system depended on the setup and the working frequency. For instance, at 300 GHz, SR is 1 mm while it is about 0.7 mm at 500 GHz.
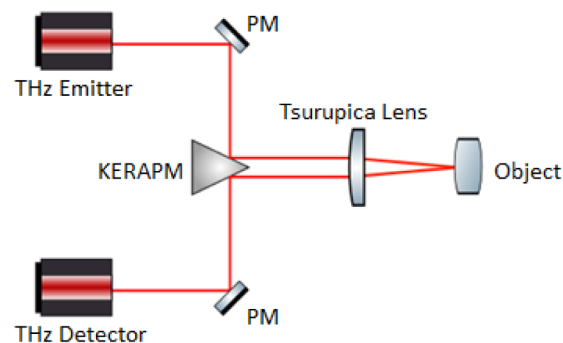

Fig. 1. Schematic view of the reflection geometry system. The THz pulses are focused onto the sapphire substrate which carries the biological sample. The maximum incident angle of pulse is θ = 10°. PM stands for planar mirror and KERAPM stands for knife-edge right-angle prism mirror.

Through their propagation, emitted pulses reflect from two dielectric interfaces. The first interface between air and sapphire gives arise to the first reflected pulse, also called lower reflection peak. The second pulse to be detected is the one reflected from the upper-interface: the sapphire-sample interface, see Fig. 2. The sample dielectric properties can be extracted from the second reflected pulse.
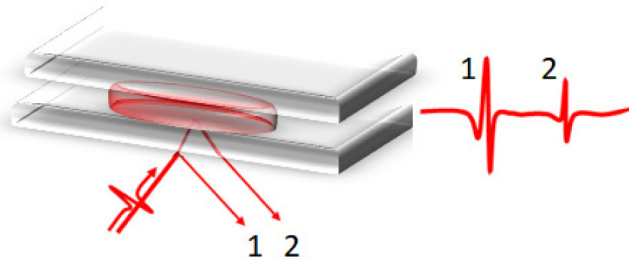

Fig. 2. Schematic representation of the biological sample (red) between the two sapphire windows (grey). The two main reflected peaks. The so-called lower reflection peak (1) arises from the air-sapphire interface while the upper reflection peak (2) is reflected from the sapphire-sample interface.

2.2 Data extraction

An essential aspect of THz time-domain spectroscopy is that both the phase and amplitude of the spectral components of pulses are measured independently. The measured amplitude and phase are directly associated to the refractive index and the absorption coefficient, thus allowing the complex permittivity extraction without Kramers-Kronig analysis. Moreover, THz spectroscopy is capable of non-destructively detecting dissimilarities thanks to the sufficiently long wavelength and the low energy of radiation that do not produce any temperature increase or photo-induced reactions within the samples. One can access the dielectric properties from the calculation of the reflection quotient which is expressed as:

$$Q(x, y, \omega) = \frac{F[E_s(x, y, t)]}{F[E_R(x, y, t)]} \tag{1}$$

where E_s stands for the sample time-domain signal with respect to its reference E_R. The dielectric properties were extracted following the procedure in *Fan et al* [39]. In particular, we employed a meticulous calibration procedure due to the extreme sensitivity to phase mismatch between the reference pulse and the sample reflection. Firstly, we ensured that the sapphire substrate was firmly held and parallel with respect to the motorized stage. Then each pixel location was measured and recorded as reference signal. Applying then the self-reference method proposed by [40], the phase misalignment, remaining the main error source and inducing dramatic consequences on extracted properties, was drastically reduced [41]. Following this, a numerical procedure was performed to calculate and to correct the baseline. Using the reference pixels, the dielectric responses of samples were calculated from each pixel signal.

2.3 Set-up calibration

Calibration was performed with liquid water since it is the main constituent of biological samples. To this aim, an averaging of 100 measurements per acquisition was performed on liquid water drops deposited onto the sapphire substrate. These measurements have been realized at different positions and the extraction was reiterated to evaluate the repeatability of the entire measurement chain. The extracted refractive index and absorption coefficient for liquid water are shown Fig. 3. The mean percentage error in the $0.2 - 1$ THz range is about 5% and 10% for the refractive index and the absorption coefficients, respectively. Beyond 1

THz, the acquired signal gives arise to the exhibited fluctuations. The air-dried injection employed in our system to remove water vapor molecules, at the hospital, was not sufficiently efficient. Consequently, the SNR drastically decreases, thus making difficult the dielectric property extractions out of the $0.2 - 1$ THz window. Our results were compared with simulations performed using a double-relaxation Debye model [42–44]:

$$\varepsilon_c = \varepsilon_\infty + \frac{\varepsilon_s - \varepsilon_1}{1 + i\omega\tau_D} + \frac{\varepsilon_1 - \varepsilon_\infty}{1 + i\omega\tau_2} \tag{2}$$

where ε_c is the dielectric function, ε_∞ stands for the dielectric constant at high frequencies, ε_s is the static dielectric constant, ε_1 is the dielectric function involved in the transition between the long and fast relaxation process, occurring over τ_1 and τ_2, respectively at pulsation ω. The double Debye theory assesses the dielectric relaxation of a certain material submitted to a given electromagnetic stimuli lying in the commonly defined THz range.

On the basis of these results, a water measurement reference was undertaken before each biological sample measurement to evaluate and to correct the impact of atmospheric variations on data extraction.
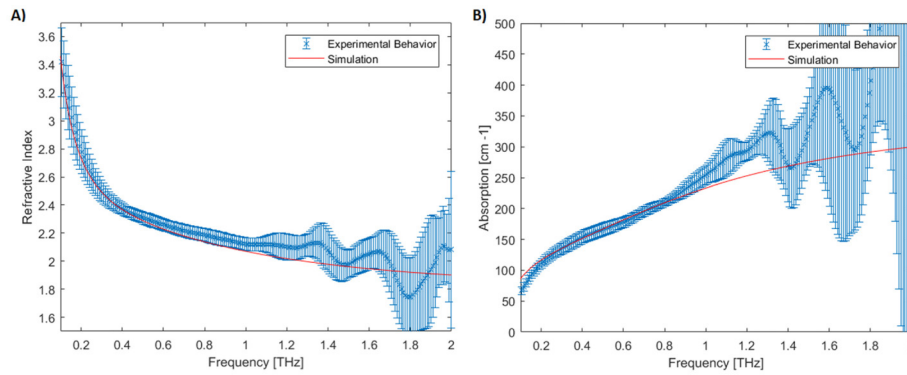

Fig. 3. A) Measured (blue) and simulated (solid red) refractive index of water. Results are obtained over 13 independent measurements. The mean percentage error between $0.2 - 1$ THz is 5%. B) Measured (blue) and simulated (solid red) absorption coefficient of water. Results are obtained from 8 independent measurements. The mean percentage error between $0.2 - 1$ THz is 10%. The values of parameters used in simulation were $\varepsilon_\infty = 3.2$, $\varepsilon_s = 78.4$, $\varepsilon_1 = 4.9$, $\tau_1 = 8.0ps$ and $\tau_2 = 0.18ps$.

2.4 Tissue registration

To correctly assess variation across and within samples, a classification of tissues regarding their biological structure is of paramount importance. The tissues have been catalogued according to the physician's diagnosis following the standard fixation and staining procedure described in section *2.5 Tissue preparation*. Samples were indexed as follows: (C), (F) and (A), denoting the presence of a cancerous fiber matrix, a healthy fiber matrix or an adipose region, respectively. A sub indexation has been established to denote the region of interest (RoI) on which we have performed a single-point measurement. Aside from particular cases, fat was always visually distinguishable. Our assumption was in a first approximation that the beam diameter variation over the $300 - 600$ GHz range was quite flat. Considering then a beam diameter about 1mm, single point measurements were performed 5 mm away from visually distinguishable adipose regions to remove fat contribution in recorded signals. The position of the THz beam within the sample is performed by superimposing an IR laser beam with the THz beam during the calibration procedure.

2.5 Tissue preparation

Sixteen freshly excised breast tissues from 11 women who underwent breast conserving surgery or total breast removal at Bergonié Institute were analyzed through the as-mentioned set-up and registration. 5 samples were invasive lobular carcinoma, 6 samples were invasive breast cancer without any specific type, 2 were metaplastic breast cancer, 1 was breast phyllodes tumor and 2 were excised from a breast reduction. Table 1 summarizes the different samples. Measurements were performed on tissues which have not been submitted to any chemical treatments. They have been immediately immersed into physiological serum after the excision surgery and carried out to the histology room to be prepared for medical diagnosis. The samples were sliced with a scalpel and a half was instantly used for THz imaging or single-point spectroscopy measurements. To ensure good contact between the measurement window and samples, tissues were slightly pressed with another sapphire substrate on the top (as in Fig. 2). Beforehand each image, line scans were processed in X and Y directions to warrant the flatness of the measurement window. A maximum of 30 minutes elapsed between the excision procedure in the operating room and the measurement, ensuring no degradation of fresh state tissues. Samples which did not match with this specific requirement have been removed from the study database. After scanning -between 45 and 60 minutes,- tissues were fixed in a formalin solution. Hematoxylin and eosin (H&E) stained section were produced to compare histological diagnosis and THz-images.

Table 1. List of the 16 freshly excised tissues analyzed. Samples having the same letter reference code belong to the same patient and have been excised simultaneously. Excision column refers to the different tissue types: adipose (A), cancerous (C) and fibers (F). Denom. stands for the denomination of the cancer type associated to its grade. Single-point measurement (SPM) and Image column refers to the experimentation done.

Sample		Excision	Denom.	Age	SPM	Image
	I	A / C / F	ILC[1] 2	74	X	X
A	II	A / C / F				
	III	A / C / F			X	
B	I	C / F	IBC[2] 3	82	X	X
	II	A / F			X	
C	I	A / C / F	MBC[3] 3	63	X	X
	II	A			X	
D	I	C / F	IBC 3	66	X	X
E	I	C / F	IBC 3	80	X	X
F	I	C / F	ILC 2	73	X	
G	I	A / C / F	IBC 2	55	X	
H	I	C / F	BPT[4] 2	63	X	
I	I	A / F	BR[5]	48	X	X
	II	A / F			X	X
J	I	A / C / F	ILC 3	82	X	X
K	I	A / C / F	IBC 3	84		X

[1]ILC: Invasive lobular carcinoma. [2]IBC: Invasive breast cancer, no specific type. [3]MBC: Metaplastic breast cancer. [4]BPT: Breast phyllodes tumor. [5]Breast reduction.

2.6 Image processing

After terahertz data acquisition and processing, the very high information contained in the data leads to new challenges for relevant feature detection. Indeed, the consistent traits for tissue recognition may be localized only on a few frequencies. For this reason, in several related works, simple parameters are selected such as peak intensity, peak to peak intensity in time domain or a selected FFT amplitude [22]. These data are then used as parameters to produce THz images. Moreover, a large number of parameters could be derived from the THz pulses and spectra based on time or frequency domain characteristics.

For our part, initial data processing involved the removal of background noise by calculating the mean of each pixel signal. Then, the mean value was removed from each pixel signal. We applied a zero-padding algorithm to the time-domain signal of each point such that only the peak arising from the second interface remains. On the time-domain information a fast-Fourier transform (FFT) is applied to convert to the frequency-domain. A frequency window is selected to clean the symmetric modulus arising from the FFT calculation and to remove the frequencies out of the 300 – 600 GHz window. We then performed PCA, a compressive technique which is an unsupervised linear algorithm traditionally used to reduce data dimensionality. It consists of an orthogonal transformation to outline a set of uncorrelated values with each successive principal component having the highest possible variance under the orthogonality constraint. The principal component function of the Statistics and Machine Learning toolbox from MATLAB was used to perform such an analysis. Close to the Karhunen-Loève transformation (KLT), PCA takes advantage from other multivariate statistics since the mathematical projection core is adapted as a function of input data. This tool completes a panel of different mathematical and/or statistical processes which can be performed for image treatments.

3. Results

3.1 Spectroscopy

As the biological architecture of breast tissue is different from one sample to the other, the light-tissue interaction is complex and difficult to physically quantify. Also, the heterogeneity of the tissue and the varying beam size with frequency are both additional complications. Nevertheless, the effective refractive index has been extracted and averaged from different tissues over several locations and is reported Table 2. Approximately 34,000 waveforms have been collected from the 16 samples analyzed. The results are shown considering the mixture of tissue presented by the measurement area and are compared with both the optical view and the histology slide. After correlation with the optical image and the stained slice, we have estimated a deviation of +/− 2 pixels with respect to the step sizes in X and Y directions. Then, we selected the pixels where there was no ambiguity about the nature of the tissue. The measurement area background may differ from the excision column mentioned in Table 1, as the mean effective refractive indices were extracted from localized tissue regions. Results indicate that in the 300 – 600 GHz range, the refractive index of C / F (cancer / fibers) regions differs from A / F (adipose / fibers) regions and pure fatty tissue. The refractive index resulting from a spectroscopic measurement over C / F regions is always higher than the one extracted for other tissue mixtures. A significant difference can be observed in the 300 – 600 GHz band. The extracted optical indices are in good agreement with published data. There is no clear evidence that the cancer type and grade play a significant role on the refractive index value. We are currently investigating a solution to propose for each sample a linear equation describing the contribution of each tissue response, on the basis of Bruggeman distribution [45]. However, results may suggest the ability of THz spectroscopy to differentiate the margins of both ILC and IBC from surrounding fibro glandular tissue and fat within the 300 – 600 GHz. A wider set of tests on ILC and IBC pathologies is however needed to draw further conclusions. Similar tendencies were also observed for assessing breast carcinoma in excised paraffin-embedded human breast tissue in both transmission and reflection imaging. Concerning metaplastic breast cancer and phyllodes tumor, they are rare cases of breast cancer. The lack of samples characterized in the THz domain, exhibiting these specific pathologies and their associated healthy tissues, does not allow a statistical conclusion. However, one should highlight that the histology slides used to correlate the THz images with tissue regions are 5 μm thick. The opto-biological interaction depth is assumed to range between 20 μm and 50μm. Thus, another approach considering the whole histology slide stack along the penetration depth would be beneficial to assess tissue heterogeneity in depth.

Table 2. Refractive index table of the freshly excised analyzed tissues. Each sample has been probed at least 3 different spatial locations. Each location has been measured 3 times in a row, with a 5 minutes time lapse between each. A spectrum averaging of 100 was used. The region of interest (RoI) refers to the tissue type mixture area we scanned. The presence of the 570 GHz water vapor absorption peak induces for some results a higher refractive index at 600 GHz than at 500 GHz.

Sample		RoI	300 ± 5 GHz	400 ± 5 GHz	500 ± 5 GHz	600 ± 5 GHz
	I	C / F	2.17 ± 0.06	2.11 ± 0.05	2.10 ± 0.02	2.14 ± 0.02
A	II	C / F	-	-	-	-
	III	C / F	2.28 ± 0.05	2.22 ± 0.06	2.21 ± 0.07	2.24 ± 0.08
B	I	C / F	2.16 ± 0.05	2.13 ± 0.05	2.09 ± 0.04	2.07 ± 0.04
	II	A / F	1.92 ± 0.01	1.92 ± 0.01	1.92 ± 0.01	1.92 ± 0.01
C	I	C / F	2.40 ± 0.05	2.29 ± 0.03	2.20 ± 0.04	2.21 ± 0.09
	II	A	1.76 ± 0.05	1.70 ± 0.04	1.66 ± 0.03	1.64 ± 0.03
D	I	C / F	2.26 ± 0.06	2.18 ± 0.02	2.12 ± 0.02	2.09 ± 0.03
E	I	C / F	2.23 ± 0.03	2.12 ± 0.04	2.08 ± 0.05	2.04 ± 0.05
F	I	C / F	2.19 ± 0.03	2.09 ± 0.03	2.00 ± 0.06	2.00 ± 0.06
G	I	C / F	2.34 ± 0.17	2.26 ± 0.13	2.20 ± 0.11	2.15 ± 0.15
H	I	C / F	2.36 ± 0.01	2.28 ± 0.01	2.24 ± 0.01	2.20 ± 0.01
I	I	A / F	1.89 ± 0.05	1.86 ± 0.04	1.85 ± 0.04	1.83 ± 0.03
	II	A / F	1.92 ± 0.06	1.87 ± 0.06	1.86 ± 0.04	1.87 ± 0.05
J	I	C / F	2.43 ± 0.03	2.34 ± 0.01	2.24 ± 0.02	2.21 ± 0.02
K	I	C / F	2.26 ± 0.06	2.13 ± 0.06	2.09 ± 0.07	2.04 ± 0.05

3.2 Image processing

We show images obtained with respect to the analysis chain described in *section 2.6*. We thus report the amplitude frequency images for the J-I and K-I samples, respectively, in Fig. 4. The amplitude of the reflected signal is directly linked to the Fresnel's coefficients at the dielectric interface. In fact, the proportion of photons reflected back into the detector is greater when the refractive index difference between the two components forming the dielectric interface is high. To interpret the following images, one has to consider the refractive index of the C-cut sapphire window being 3.10 ± 0.05. Thus, the clearest regions correspond to adipose tissue, having the lower refractive index, thus increasing the proportion of pulses reflected back into the detector. On the other hand, the darker surfaces are induced by tissues having a higher refractive index, closer to the one of the sapphire window, *id est* fibers and cancer. Since it is tedious to compare different samples between each other, the main relevant parameter is the contrast revealed within each individual sample. Different areas can be segmented and labelled and some interesting correlations between THz-images and histology slides can be found.

On reported J-I sample imaging Fig. 4 (top), frequency images exhibit some structures in the 300 – 600 GHz range. More specifically at 300 and 400 GHz, respectively. Interesting demarcations between cancer and fibers are reported. However, some malignant sections are not well delineated through the images. Dashed regions are added to bring into light a good correlation between the histology slide and the THz-images. Considering the external boarder of the unhealthy regions, we observe a quite good agreement between adipose and the fiber matrix, containing a substantial amount of cancer cells. The second sample imaging reported here in Fig. 4 (down) suffers from the tissue structure (*i.e.* topology) which induces a non-perfect contact with the sapphire window, leading to missing area and artifacts. Consequently, in that case the image contrast variation mostly denotes the sample topology instead of its biological composition.
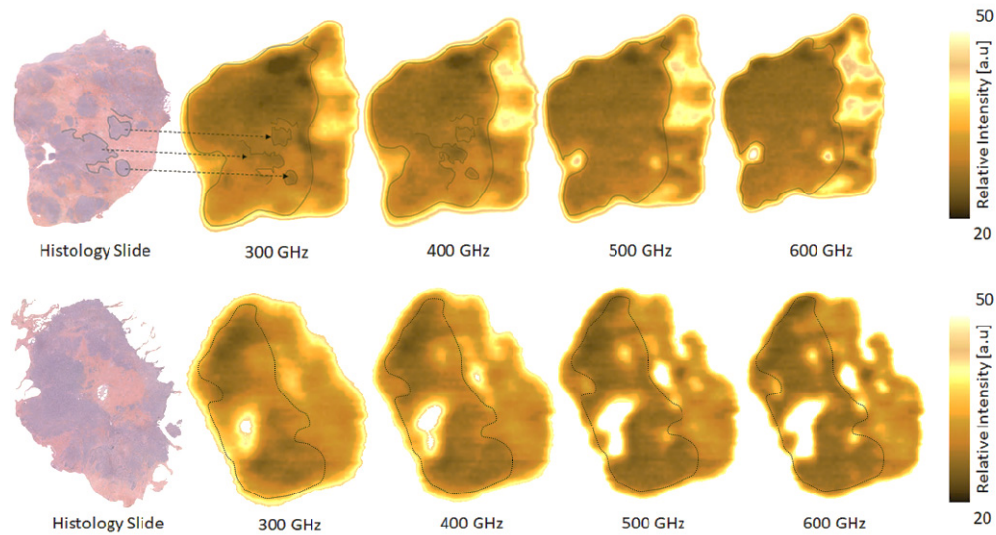
Fig. 4. Comparison between histology slides and corresponding THz-images at 300, 400, 500 and 600 GHz for the J-I (top) and K-I (down) samples, respectively. Adipose is purged out during paraffinization and it is then not visible in histology sections. On J-I sample (top) some interesting features between the histology slide and the THz-images are highlighted through dashed regions. Frequencies ranging from 300 GHz to 400 GHz exhibit the most intriguing demarcations between cancerous regions and healthy fibers. K-I sample images (down) suffer from the tissue topology, presenting several hills and valleys.

An additional set of images recorded at 500 GHz are reported Fig. 5. Tumor areas determined on the images are in good agreement with histology slides. Contrast mechanisms are still under investigation, but there is a substantial difference between healthy breast tissue and carcinoma in the BiCMOS compatible frequency range. One of the reasons for these results could be an increase in the vasculature associated with faster cell division and higher cell densities. These important physiologic changes generally lead to an increase of the water content -to which THz imaging is particularly sensitive- and to a decrease of the lipid concentration compared with healthy tissue. Moreover, lipids have]ing low absorption within the THz range, the water content increase should lead to significant contrast. However, water concentration is likely not the only factor responsible for contrast. Numerous papers demonstrated that biologic samples that were fixed in formalin, dehydrated and embedded in wax for histopathologic examination, present significant contrast between tumor and the surrounding healthy tissues. Other possibilities linked to cell density or the presence of specific proteins may also be responsible for contrast.
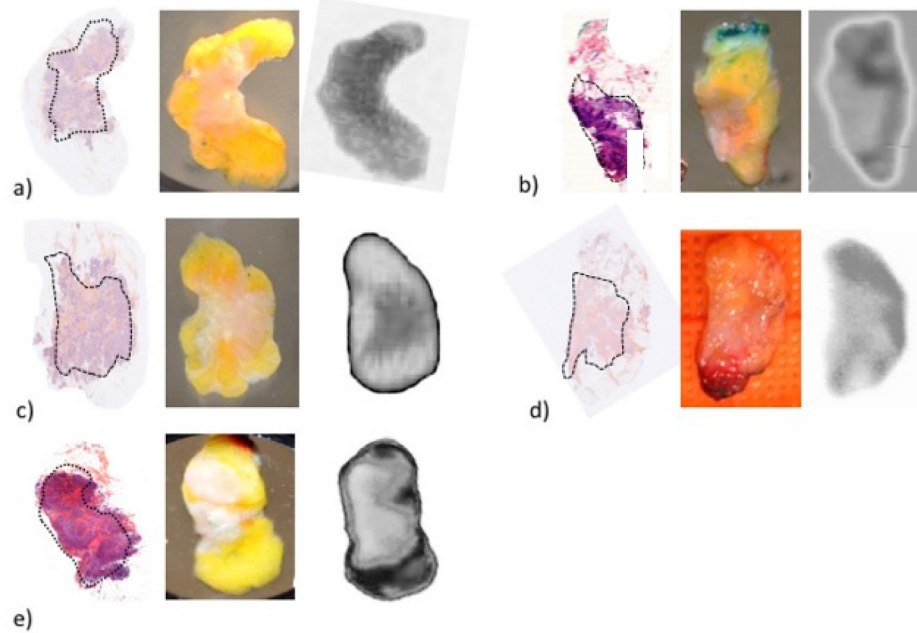
Fig. 5. Additional set of samples THz-imaged. For each sample, from left to right: 5 µm stained tissue slice; optical photograph; THz-image at 500 GHz. Each section contains a healthy region and a malignant one (dashed area on histology slides).

To enlarge the field of inquiry, principal component analysis was performed on the amplitude of the FFT spectrum of I-II sample, a pure healthy tissue from a breast reduction surgery, to highlight its capability to distinguish between fat and fibers. As a first procedure, we have inserted the entire pixel collection as variables into the PCA model. Thus, sapphire/air dielectric interface pixels have also been taken into account. The first principal component image depicted in Fig. 6 exhibits interesting features which can be interpreted as the sample biological structure. Adipose and fibrous regions, distinguishable on the visible image are well defined. Moreover, more complex structures are exhibited where pixels do not strictly belong to a unique tissue type. On the basis of such an observation a method based on the pixel value comparison with its surrounding could be investigated to establish the first numerical definition of tissue kind segmentation. Such an accomplishment would be of paramount importance to assess tumor margins. The second component shows less contrast within the tissue than the first component. However, the boarders of the tissue are well delimitated. These two observations on both the first and the second components are interesting since the first component could be used to numerically segment within the sample while the second component could serve as a boarder indicator. Clearly, deeper investigations are needed to statistically consider PCA as a tool for tissue segmentation. Moreover, demarcation within tissues exhibiting malignant zones would be even more difficult. Nevertheless, PCA should be extensively studied to highlight its role in tissue segmentation.
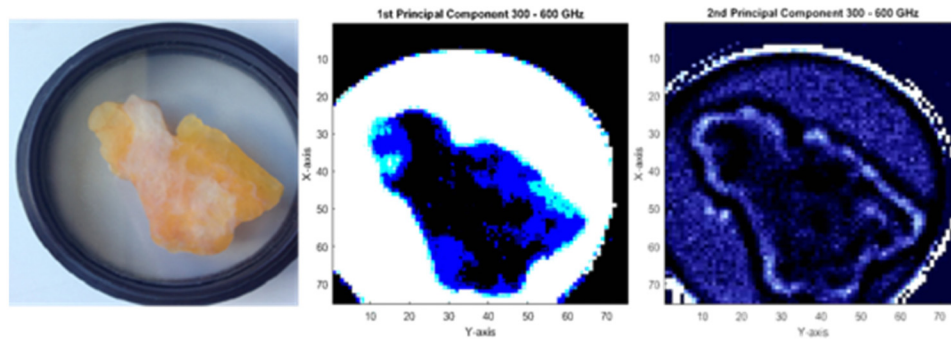
Fig. 6. Principal component analysis over the entire FFT amplitude image. From left to right: visible view of I-II sample; first principal component image within the 300 – 600 GHz band; second principal component image within the 300 – 600 GHz band.

A second investigation using PCA, focusing only on tissue pixels has been performed. Results are reported through Fig. 7. A specific location of the tissue exhibiting clearly the two observable tissue kinds on the I-II sample has been selected. The selected area consists of 182 different pixels. The score plot of the two orthonormal first components has clearly classified the pixels in two distinct groups. Following this, a check procedure was performed to determine which kind of tissue each group belonged. Based on the frequency amplitude, we have attributed each group of pixels to a specific tissue type. However, some pixels do not strictly belong to one particular tissue. One explanation would be to suggest that these pixels consist of a fat/fiber distribution or some outliers.
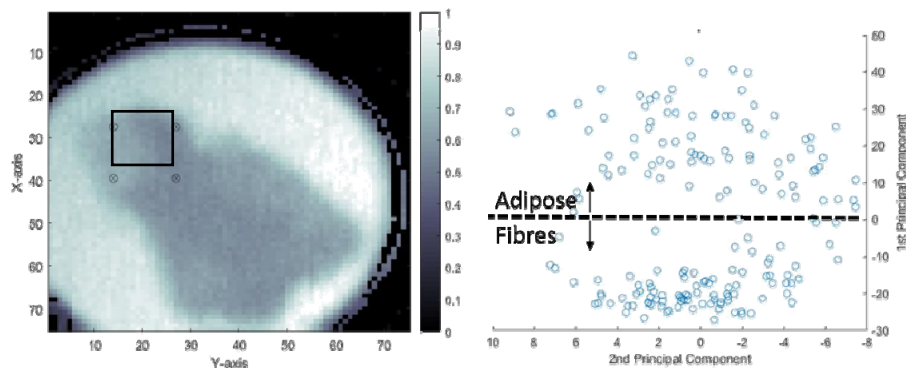

Fig. 7. From left to right: FFT amplitude at 390 GHz. The squared area is the region over which the PCA has been performed; score plot of PC1 and PC2, for the 182 pixels contained in the PCA model. A clustering between adipose and fibrous tissue is delineated by a dashed line.

4. Conclusion

In this pilot study we focused on the ability of the under-exploited 300 – 600 GHz frequency band to differentiate between breast tissue types. To do so, a far-field reflection geometry THz imaging system was situated in the Department of Pathology at Bergonié Institute, the oncology laboratory partner. Sixrwwn freshly excised breast tissue samples have been collected and analyzed in view of their responses within the 300 – 600 GHz narrow frequency window. The refractive index of these tissues was calculated and compared between each other at several frequencies. Our results confirm the tendency for tissue matrices containing malignant cells to exhibit a higher mean refractive index than the ones consisting of healthy tissues. Such a difference is observed over the studied frequency band. THz-images have also been produced; in the majority of them, a quite good optical correlation between histology

slides and THz-delimitated area is observed. However in order to increase the efficiency of such measurements, a THz near-field imaging matrix array is under study and will be deployed to assess, with greater resolution, higher speed and lower cost breast tissue classification.

Principal component analysis was conducted on a healthy tissue provided from a breast reduction, to manage the early stage of development of an automatic tool, classifying tissue types with no input from humans. A clear demarcation within the tissue between adipose and fibers has been seen through the first principal component image. The second component extracted through the analysis gave arise to an image exhibiting interesting delimitations between the sample boarders and the sapphire substrate. A PCA focused on the tissue structure successfully classified pixels in two distinct groups, respectively belonging to either adipose or fibrous breast tissue. A few pixels were not clearly attributed to a specific group and we will work on the improvement of data separation.

Ethical approval

Human tissue analysis and measurements have been performed in view of the fundamental ethical stipulated in the Helsinki declaration and its later revisions. Samples were obtained with the written approval of each patient undergoing an excision procedure.

Funding

Acknowledgments

Disclosure

The authors declare that there are no conflicts of interest related to this article.